P 419

Prediction of Porosity and Sand Fraction from Well Log Data using ANN and ANFIS: a comparative study

Soumi Chaki[1*], Akhilesh K. Verma[2], Aurobinda Routray[1], Mamata Jenamani[3], William K. Mohanty[2], P. K. Chaudhuri[4], S.K. Das[5]

Summary

Reservoir characterization is a difficult problem due to nonlinear and heterogeneous physical properties of the subsurface. In this context, we present a case study to compare Artificial Neural Network (ANN) with Adaptive Neuro Fuzzy Inference System (ANFIS) for predicting two reservoir characteristics: porosity and sand fraction from well log data. The predictor variables are gamma ray content (GR), density (RHOB), P-sonic (DT), and neutron porosity (NPHI) logs. We use the data from a hydrocarbon field located in western part of India. It has been shown that while prediction results from both the models are comparable in terms of performance evaluators such as root mean square error, correlation coefficient etc., ANFIS has an inherent edge over ANN to describe and model uncertainties; whereas the computational complexity is high for ANFIS, making ANN a natural choice for the dataset under consideration. The research concludes that choice of the prediction model is closely associated with the nature of the dataset, and newer and complex methodologies need not always perform better in terms of accuracy of the result and computational complexity.

Keywords: *Reservoir characterization, well log data, Artificial Neural Network (ANN), Adaptive Neuro Fuzzy Inference System (ANFIS), Blind Prediction*

Introduction

Reservoir characterization - the problem of finding a good quality reservoir, includes the task of predicting geophysical parameters such as porosity, sand fraction and permeability. Porosity is a key indicator associated with potential hydrocarbon volume, whereas sand fraction value is a parameter to indicate porosity. Though, sand fraction and porosity are interrelated; they are many times interpreted separately. Therefore, these two properties are needed to be predicted independently before making any conclusive decision about the field productivity. Since there is no direct measurement for these parameters, they are to be computed from other geophysical logs (Hamada and Elshafei, 2010) or seismic attributes (Hou et al., 2008). This process also requires repeated intervention of experts for fine tuning the results. Standard regression methods are not suitable for this problem due to high degree of unknown nonlinearity. The problem is further complicated because of uncertainties associated with lithological units. In this context, ANN and its variants with Fuzzy Logic are considered to be useful tools to establish the mapping between lithological and well log properties (Rogers et al., 1992; Lim, 2005; Aminian et al., 2005; Kaydani et al., 2012).

Estimation of petrophysical properties is associated with a number of complex tasks such as data fusion (i.e. integration of data from various sources), data mining (i.e. information retrieval after analyzing those data), formulation of knowledge base, and handling of uncertainty. The applications of advanced statistical, machine learning and pattern recognition techniques to such problems have received considerable interest amongst the researchers in oil-gas sector (Nikravesh, 2004; Nikravesh et al., 2001a). The objective of these types of studies is to identify potential zone for drilling a new well (Nikravesh et al., 2001b). In this research

[1]Department of Electrical Engineering, IIT Kharagpur, [2]Department of Geology and Geophysics, IIT Kharagpur
[3]Department of Industrial Engineering and Management, IIT Kharagpur, [4]INTEG, GEOPIC, ONGC, Dehradun
[5]ONGC, Baroda

Department of Electrical Engineering, IIT Kharagpur, Kharagpur- 721302
Email address: soumibesu2008@gmail.com

domain, different techniques such as type-2 Fuzzy Logic system (Olatunji et al., 2011), and hybrid systems (Helmy et al., 2010; Anifowose et al., 2011) are also applied to estimate petrophysical properties.

In the present study, two methods, Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) are applied to predict porosity and sand fraction values from gamma ray, density, P-sonic and neutron porosity logs. The data used in this study is collected from a set of four well logs located at a hydrocarbon field in western onshore of India. Most of the authors combine the entire data into one set and then randomize it to generate the training and testing sets in similar studies. For example, 70% of the total available data may be used as training set and remaining 30% as testing set. In this study, instead of following this mechanism, we use the combined data from three wells as training set and blindly predicted the porosity and sand fraction value of the fourth well. We setup the right network structure for both the models by conducting large number of experiments. Several iterations are then made on the selected network and mean of the runs is considered to ensure integrity of the results obtained. We observe that prediction results from both the models are comparable in terms of performance evaluators such as root mean square error, correlation coefficient etc. However, the computational complexity is high for ANFIS, making ANN a natural choice for the dataset under consideration. Though, earlier researches in this field mostly report the advanced techniques such as ANFIS to be better, we demonstrate that choice of the prediction model is closely associated with the nature of the dataset, and advanced methodologies need not always perform better in terms of accuracy of the result and computational complexity.

The rest of the paper is structured as follows: we first describe the data used in this study; next an overview of the two methodologies (ANN and ANFIS) is provided; a brief description on the performance evaluators used in this work is given in the section following. Then, experimental results are reported. Finally, the paper concludes with the discussion of the results and a brief overview of the future scope.

The data

In this study, data from four closely spaced wells located in the western onshore of India are used. Henceforth, these wells are to be referred as A, B, C, and D respectively. The predictor variables used in this study are gamma ray content (GR), density (RHOB), P-sonic (DT), and neutron porosity (NPHI). Gamma ray log represents gamma radiation of different formations in American Petroleum Institute (API) unit along the depth. Depending on the changes in mineralogy and porosity, density differs among various rock types. It is represented by grams per cubic centimeter unit. P-sonic is an acoustic log which represents travel time of P-waves versus depth in micro second per feet. Neutron porosity is expressed in terms of per unit and attuned to read the true porosity assuming that the pores are filled with fresh water. The geophysical properties e.g. sand fraction and porosity are modelled using these four predictor variables.

The Methodology

The theories associated with Artificial Neural Network (ANN) and Adaptive Neuro-fuzzy Inference System (ANFIS) and its application framework in the context of the case study are briefly discussed in this section.

A. Feed-forward Artificial Neural Network Approach (Haykin, 1999)

In this study, a multilayer feed forward network is selected as a standard procedure. The parameters of the network are initialized heuristically and then adjusted experimentally keeping the improvement of the training and testing performances in view. The methodology followed in training and testing of ANN is shown in Figure 1 and the corresponding network structure is depicted in Figure 2. The structures of input and output layers consist of four and two neurons respectively. There is only one hidden layer. The number of hidden neurons is adjusted to improve the training and testing results while keeping the number of available training patterns in mind. The total number of weights is significantly less than the number of available training patterns to avoid the problem of over fitting.

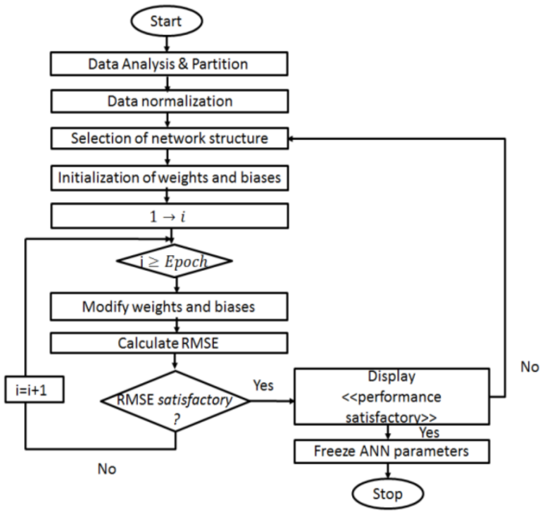

Figure 1: Flow chart of the methodology using ANN

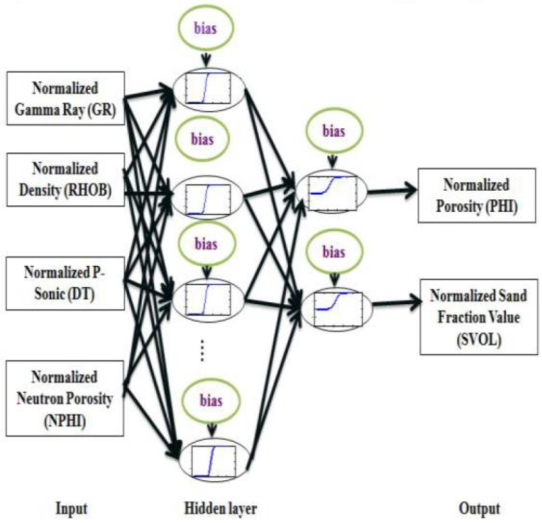

Figure 2: Network Structure used in ANN training

In the hidden layer, hyperbolic tangent sigmoid transfer function is used. The tangent sigmoid transfer function is an automatic choice for researchers to use in hidden layer to achieve the bi-directional swing (Kalman et al., 1992; Leshno et al., 1993). Moreover, the learning rate of the network is faster when the network is anti-symmetric. On the other hand, generally, the activation function used in the output layer is non-symmetric. Therefore, log-sigmoid transfer function is used in the output layer. The target variable is normalized within the range of output activation function keeping some offset from limiting value of the activation function. Otherwise, the back propagation algorithm inclines to drive the free network parameters to infinity. As a result, the learning process will slow down (Haykin, 1999). Hence, the target variables are normalized between 0.1 and 0.9 (Equation 1) so that it does not fall into the saturation region of the log-sigmoid function.

$$PHI_{normalized}=a+\frac{PHI-PHI_{min}}{PHI_{max}-PHI_{min}}*(b-a) \quad (1)$$

where a and b are two constants representing lower and upper limit of that range. The predictors e.g. gamma ray, density etc. are normalized as:

$$GR_{normalized}=\frac{GR-GR_{mean}}{GR_{standard\ deviation}} \quad (2)$$

The available data is partitioned into training and testing sets. Here, combined dataset from three wells are used for training and the trained network is tested using the remaining well log. Scaled-conjugate-gradient-backpropagation algorithm is used to train the network as it requires no user-dependent parameters and the speed of convergence is very high. Starting with a very simple structure, a network with 75 neurons in the hidden layer and two neurons in the output layer has been finally chosen. The network is trained using 1000 number of epochs keeping minimum gradient and goal constant at 0.00001 and 0.0001 respectively.

B. Adaptive Neuro-fuzzy Inference System approach (Jang, 1993)

Fuzzy logic is an extension of conventional Boolean logic (zeros and ones) developed to handle the concept of "partial truth" values between "completely true" and "completely false". A fuzzy set consists of two components: element and membership value. A fuzzy set $A(x)$ is represented as:

$$A(x)=\{(x,\mu A(x)),x\in X\} \quad (3)$$

where X is the universal set, x is an element and μA is membership function.

Fuzzy Inference System (FIS) plays a crucial role in NeuroFuzzy system. This inference mechanism is carried out by a set of rules operating upon variables which are fuzzified through various membership functions. The number and type of fuzzy rules, membership functions and the method of inference are the basic building blocks of the FIS system. Fuzzy if-then rules are applied to detect the inexact modes of reasoning that play an significant role in the human ability to take decisions in an environment of uncertainty and imprecision. Here, we have used an Adaptive NeuroFuzzy Inference System (ANFIS) to model porosity and sand fraction from predictor variables. Figure 3 describes the workflow of ANFIS. Firstly, dataset is divided into training and testing patterns and normalization is carried out.

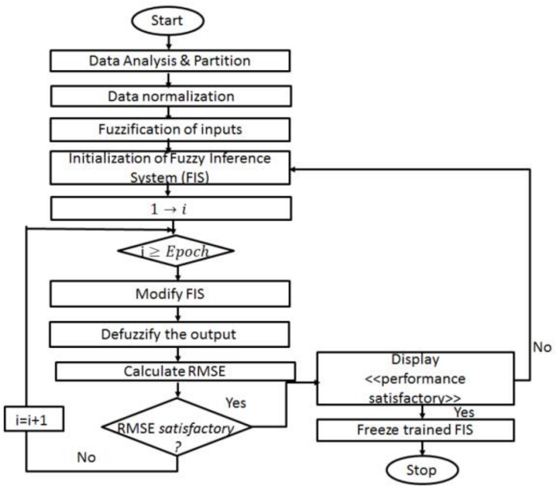

Figure 3: Flow chart of ANFIS system

The selected membership function to fuzzify predictor variables is a generalized bell shaped. The function can be expressed by the following equation:

$$f(x;a,b,c) = \frac{1}{1 + \left|\frac{x-c}{a}\right|^{2b}} \tag{4}$$

where c represents the center of the corresponding membership function, a represents half of the width; a and b together determine the slopes at the crossover points. The function is continuous and infinitely differentiable. Therefore, generalized bell shaped membership function is selected as learning may go from +infinity to –infinity. Next, rule base is generated from the normalized dataset. The initialized Sugeno type FIS is trained using a combination of least squares estimation with back propagation. Finally, the trained FIS validated using testing dataset. The performance of the trained system is evaluated in terms of correlation coefficient, root mean square error and CPU execution time.

Performance Evaluators

We use two quantitative measures to judge the accuracy of the proposed methods - correlation coefficient (CC) and root mean square error (RMSE). The CPU execution time is used to compare the computational complexity of both the algorithms. The formula of calculating correlation coefficient (CC) can be expressed as:

$$CC = \frac{\sum (x-x')(y-y')}{\sqrt{\sum (x-x')^2 \sum (y-y')^2}} \tag{5}$$

where x and y represent actual and predicted value of a variable, whereas x' and y' represent the mean of the actual and predicted values.

Root mean square error (RMSE) can be calculated as follows:

$$\text{RMSE} = \sqrt{\frac{(x_1-y_1)^2 + (x_2-y_2)^2 + \ldots + (x_n-y_n)^2}{n}} \tag{6}$$

where x_n, y_n are the actual and predicted variables and n is the size of the data..

Results and Discussion

The prediction models are trained using the combined data of the three well logs. Then, the trained network structure is validated using the data of the fourth well which was not part of the training set. Fig. 4(a)-(c) represent the cross plots between actual and predicted porosity value by ANN and ANFIS respectively for blind prediction of Well A. Similarly, the cross plots between actual and estimated sand fraction values are depicted in Fig. 4(b)-(d) for the same well. The red straight line present in each of the plots in Fig. 4 signifies linear fitting of the cross plot. The correlation coefficients for both porosity and sand fraction for all wells are tabulated in Table I and II respectively. The high values of correlation indicate that both the models are working effectively in estimating lithological properties from given predictors (i.e. input well logs). The second criterion for performance evaluation i.e. root mean square error is calculated in training and testing phase for four wells and reported in Table III and Table IV. As evident from the tables both the methods give equivalent results. Both the models are run on the Intel(R), 2.30 GHz machine with 4 GB RAM. Table V shows time taken by both the models. As shown, execution time for the ANFIS is around four fold in comparison to the ANN model.

TABLE I
CORRELATION COEFFICIENT OF POROSITY PREDICTION OF FOUR WELLS BY ANN AND ANFIS

Well Number	ANN		ANFIS	
	Training	Testing	Training	Testing
A	0.9954	0.9176	0.9910	0.9211
B	0.9797	0.9452	0.9683	0.9344
C	0.9771	0.9708	0.9687	0.9603
D	0.9734	0.9903	0.9569	0.9839

TABLE II
CORRELATION COEFFICIENT OF SAND FRACTION PREDICTION OF FOUR WELLS BY ANN AND ANFIS

Well Number	ANN		ANFIS	
	Training	Testing	Training	Testing
A	0.9992	0.9710	0.9986	0.9712
B	0.9844	0.9763	0.9801	0.9729
C	0.9849	0.9802	0.9792	0.9826
D	0.9814	0.9959	0.9762	0.9948

TABLE III
ROOT MEAN SQUARE ERROR OF POROSITY PREDICTION OF FOUR WELLS BY ANN AND ANFIS

Well Number	ANN		ANFIS	
	Training	Testing	Training	Testing
A	0.0171	0.1483	0.0078	0.0462
B	0.0381	0.0644	0.0154	0.0233
C	0.0397	0.0495	0.0169	0.0166
D	0.0446	0.0284	0.0180	0.0177

TABLE IV
ROOT MEAN SQUARE ERROR OF SAND FRACTION PREDICTION OF FOUR WELLS BY ANN AND ANFIS

Well Number	ANN		ANFIS	
	Training	Testing	Training	Testing
A	0.0084	0.1350	0.0177	0.1458
B	0.0411	0.0567	0.0479	0.0636
C	0.0399	0.0506	0.0494	0.0513
D	0.0448	0.0283	0.0536	0.0242

TABLE V
CPU EXECUTION TIME OF POROSITY PREDICTION OF FOUR WELLS BY ANN AND ANFIS

Well Number	ANN	ANFIS
	Training and testing (seconds)	Training and testing (seconds)
A	82.94	357.88
B	73.54	334.45
C	80.76	340.02
D	81.96	345.27

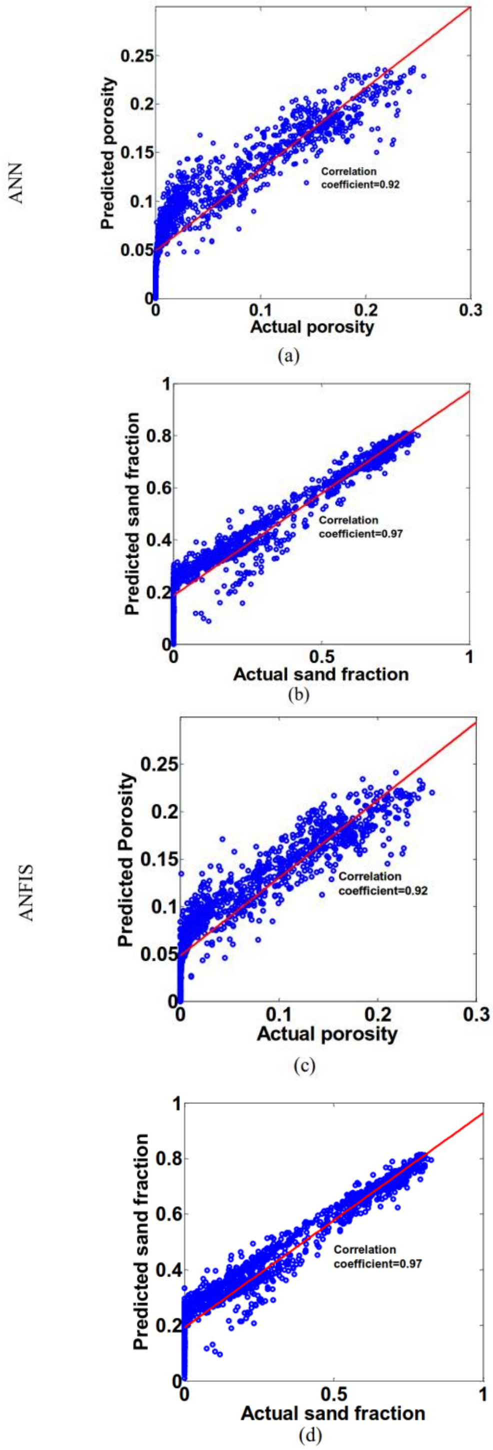

Figure 4: Cross plots between target (actual) and predicted variables by ANN and ANFIS for Well A (a)-(c)porosity, (b)-(d) sand fraction

To sum it up, the ANN outperformed the ANFIS model in terms of execution time parameter while both give almost comparable results in terms of correlation coefficient and root mean square error.

Conclusions and Future scopes

In this study, sand fraction and porosity values are estimated from gamma ray content (GR), density (RHOB), P-sonic log (DT), and neutron porosity (NPHI). Though ANFIS is a relatively recent method and claimed to be advantageous over ANN in several literatures; in this case study, ANN is established as a better prediction model in terms of comparable correlation coefficients, root mean square error, and lower execution time.

The contribution of this paper can be listed as follows: Firstly, the approach of validating the trained models through blind prediction is more practical than the common approach of dividing the available dataset into training and testing patterns combining all the data together. Secondly, ANN is proved to be better prediction model compared to ANFIS in terms performance evaluators which is a contradiction with the earlier results in the proposed field of research. Finally, the selection of transfer and membership functions is made based on the logical conclusions on the data description of the input and output variables which has been rarely adopted or at least reported in the earlier works.

The work can be extended to build a prediction model integrating seismic data and limited number of borehole data to produce a map representing values of certain petrophysical properties along a whole study area. Inclusion of seismic data in the prediction model would help to predict petrophysical properties away from the boreholes. Moreover, a rule based model which would identify the potential zone for drilling new wells can be built integrating numerical data and domain knowledge.

Therefore, it can be concluded that the selection of a particular technique depends on the dataset concerned and should be judiciously selected considering the tradeoff between the computational complexity and accuracy of the result.

References

Hamada, G.M. and Elshafei, M.A., 2010, Neural network prediction of porosity and permeability of heterogeneous gas sand reservoirs using NMR and conventional logs, NAFTA, **61** (10), 451– 460.

Hou, J., Takahashi1, T., Katoh1, A., Jaroonsitha, S., Chumsena, K. P. and Nakayama, K., 2008, Application of seismic attributes and neural network for sand probability prediction — A case study in the North Malay Basin; Bulletin of the Geological Society of Malaysia, 54, 115–121, doi: 10.7186/bgsm2008018.

Rogers, S. J., Fang, J. H., Karr, C. L., and Stanley, D. A., May, 1992, Determination of lithology from Well logs Using a Neural Network, AAPG Bulletin, 76, 731-739.

Lim, J. S., 2005, Reservoir properties determination using fuzzy logic and neural networks from well data in offshore Korea, J. of Petroleum Science and Engineering, 49, 182–192.

Aminian, K. and Ameri, S., December, 2005, Application of artificial neural networks for reservoir characterization with limited data, J. of Petroleum Science and Engineering, 49, 212-222.

Kaydani, H., Mohebbi, A., and Baghaie, A., 2012, Neural Fuzzy System Development for the Prediction of Permeability From Wireline Data Based on Fuzzy Clustering, Petroleum Science and Technology, 30, 20362045.

Nikravesh, M., 2004, Soft computing-based computational intelligent for reservoir characterization, Expert System with Application, 26, 19-38.

Nikravesh, M., Adams, R. D., and Levey, R.A., 2001a, Soft computing: tools for intelligent reservoir characterization (IRESC) and optimum well placement (OWP), J. of Petroleum Science and Engineering, 29, 239–262.

Nikravesh, M., and Aminzadeh, F., November, 2001b, Past, present and future intelligent reservoir characterization trends, J. of Petroleum Science and Engineering, 31, 67-79.

Olatunji, S.O., Selamat, A., and Abdulraheem, A., 2011, Modeling the permeability of carbonate reservoir using type-2 fuzzy logic systems, Computers in Industry, 62, 147163.

Helmy, T., Anifowose, F., and Faisal, K., 2010, Hybrid computational models for the characterization of oil and gas reservoirs, Expert systems with Applications, 37, 53535363.

Anifowose, F. and Abdulraheem, A., July, 2011, Fuzzy logic-driven and SVM-driven hybrid computational intelligence models applied to oil and gas reservoir characterization, J. of Natural Gas Science and Engineering, 3, 505-517.

Haykin, S. S., 1999, Neural Networks: A Comprehensive Foundation. Prentice Hall.

Kalman, B. L., and Kwasny, S. C., 1992, Why Tanh: choosing a sigmoidal function, International Joint Conference on Neural Networks, 4, 578-581.

Leshno, M., Lin, V. Y., Pinkus, A., and Schocken, S., 1993, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural Networks, 6, 861-867.

Jang, J. S. R., May, 1993, ANFIS: Adaptive-NetworkBased Fuzzy Inference System, IEEE Transaction on System, Man and Cybernetics, 23, 665-685.